DOCUMENT RESUME

ED 303 259 PS 017 770

AUTHOR Langford, Peter E.; Claydon, Leslie F.
TITLE The Development of Justifications for Moral Judgements.
PUB DATE 88
NOTE 33p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Adolescents; *Children; Foreign Countries; Individual Development; *Moral Values; *Research Methodology; Social Influences; *Value Judgment
IDENTIFIERS *Australia; *Coding; Developmental Theory; Kohlberg (Lawrence)

ABSTRACT

This document reports on two studies that used a novel approach to categorize justifications for moral judgments made by Australian subjects from 7 to 15 and from 7 to 21 years of age. The coding categories consisted of levels of abstraction and reflection. Level of abstraction was determined by the reference group used in justifications. Levels of reflection were descriptive (rules and rights), naturalistic (normative action or belief), and theoretical (theoretical reasons for rules). The first study showed that some developmental trends suggested by earlier studies could be replicated, while others revealed by the new coding system were established. These are interpreted as supporting a social focus rather than a cognitive limitation view of the growth of moral reasoning. There was empirical support for earlier suggestions that there are appreciable differences in the balance of justification types elicited by different topics within the domain of familiar moral judgments. The second study indicated that Colby and Kohlberg's method of categorizing moral justifications confounds form of justification with content of judgment and rule. When these two aspects are clearly separated, the indexing of growth toward moral autonomy is more accurate. Use of the suggested scoring method also overcomes problems that arise from the lack of clear conceptual definition of the face validity of Kohlberg's stages. A bibliography is provided. (RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED303259

The Development of Justifications for Moral Judgements

Peter E. Langford and Leslie F. Claydon

School of Education, La Trobe University, Bundoora 3083, Victoria, Australia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Peter E. Langford
Leslie F. Claydon

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

PS 021720

BEST COPY AVAILABLE

The Development of Justifications for Moral Judgements

Abstract

Two studies are reported using a novel approach to categorising justifications for moral judgements by subjects from 7 to 15 and from 7 to 21 years respectively. The first study showed that some developmental trends suggested by earlier studies could be replicated, while others revealed by the new coding system were established. These are interpreted as supporting a social focus rather than a cognitive limitation view of the growth of moral reasoning. Earlier suggestions, previously unsupported by clear empirical evidence, that there are appreciable differences in the balance of justification types elicited by different topics within the domain of familiar moral judgements, were given empirical support. The second study showed that Colby and Kohlberg's method of categorising moral justifications confounds form of justification with content of judgement and rule. It is suggested that growth towards moral autonomy is more accurately indexed when these two aspects are clearly separated. Use of the suggested scoring method also overcomes problems with the face validity of Kohlberg's stages arising from their lack of clear conceptual definition.

The Development of Justifications for Moral Judgements

Kohlberg's view of the development of justifications for moral judgements rests on evidence about the validity and reliability of his successive methods of scoring moral judgement interviews. Although the details of the scoring procedure have undergone continual modification up to Colby and Kohlberg (1987b), it will be assumed in this discussion that this most recent method bears enough similarity to earlier versions of standard issue scoring to warrant comparison with those earlier versions. Some comments will also be made about results obtained using previous versions of Kohlberg's scoring techniques.

The reliability of standard issue scoring has generally been found to be quite high in relation to test-retest, alternate form and inter-rater reliability, at least within Kohlberg's own Moral Judgement Interview (see Cortese, 1984; Colby and Kohlberg, 1987a).

We can consider two kinds of face validity in relation to Kohlberg's methods of assessing moral justification. The first is called ecological validity by Cortese (1984) and refers to the extent to which Kohlberg's dilemma situations adequately sample real life situations. Simpson (1974), Yussen (1977), Siegal (1980) and Gilligan (1982) have all criticised Kohlberg's dilemmas for lacking ecological validity in this sense. However, while the dilemmas in Kohlberg's Moral Judgement Interview lack such validity, other dilemmas and interviews do produce a similar pattern and apparent sequence of stages, though the alternate form reliability obtained with such interviews is less than within the Moral Judgement Interview (Davison, 1979; Gibbs et al. ,1982; Rest, 1983).

Moral Judgement Interview (Davison, 1979; Gibbs et al. ,1982; Rest, 1983).

The second kind of face validity concerns the relation between the definitions of stages and the huge number of detailed subcategories (over 700 in the 1987b version) into which standard issue scoring divides justifications. Critics of this and earlier versions of Kohlberg's scoring methods have found it hard in many cases to see why a given type of justification is assigned to a particular stage (e.g. Simpson, 1974; Trainer, 1977; Crittenden, 1978; Siegal, 1980; Cortese, 1984). Such concerns have not been adequately answered by the later development of standard issue scoring when viewed from the standpoint of face validity. Such scoring still does not use overall definitions to asssign justifications to stages, but instead tells the rater which of the huge number of initial categories, called criterion judgements, a reply belongs in and then assigns that category to a stage. It is often hard to understand the reasons given for assigning criterion judgements to stages.

Colby and Kohlberg (1987a) argue that the two most important kinds of construct validity possessed by standard issue scoring are invariance of stage sequence and that subjects tend to produce the same stage usage across different dilemmas as well as within individual dilemmas (generality of stage usage). Unlike earlier findings using sentence and story rating (Kohlberg and Kramer, 1969), evidence from longitudinal studies using standard issue scoring, including the most recent that use intermediate half stages between the major stages, have been supportive of the invariance claim (Cortese, 1984; Colby and Kohlberg, 1987a). Kuhn (1976) and Cortese (1984) have claimed that this invariance evidence is

flawed because longitudinal evidence that shows either no change in average stage level or any increase is considered confirmatory, while only stage regression is disconfirmatory. This seems, however, to neglect the consistent pattern of shifts in the balance between stages that occurs in such studies, which does provide apparent support for Kohlberg's position.

Evidence for generality can be summarised under two heads: evidence that subjects show similar stage scores across interviews and evidence that subjects score mainly at one or two adjacent stages within an interview. Rest (1983) has pointed out that the dilemmas ultimately included in the Moral Judgement Interview were specially selected to produce high generality of stage usage across interviews and across dilemmas. Studies by Davison (1979) and Gibbs et al. (1982) show that for other dilemmas such generality is considerably lower. Cortese (1984) also points out that as most subjects in generality studies have been at stages 2 or 3, generality is artificially inflated because of the small variation in subjects' scores.

Colby and Kohlberg (1987a, p. 115) consider that the finding that nearly all subjects score mostly in one or two adjacent stages supports the claim that generality exists in the second sense. Findings, from the DIT and from comprehension measures, that stage use is spread more widely are met by the claim that production provides a better assessment of a subject's reasoning than comprehension or choice. However, Colby and Kohlberg (1987a, p. 90) artificially decrease the spread of a subject's scores across stages in their production studies by assuming that for any stage for which a subject scores less than 10 percent, such

scores are due to error. Thus evidence for this kind of generality is also doubtful.

Kohlberg vacillated in his attitude to attempts to establish construct validity by prediction of subjects' moral behavior. While Scortese (1984) reports Kohlberg as being concerned to improve such prediction and Broughton (1975) reviewed a number of studies that did find such evidence and considered this as relevant, Colby and Kohlberg (1987a, chap. 2) are definite that it is not their aim to seek such support. Notwithstanding this, it is of interest that both Blasi (1980) and Cortese (1984) concluded that the strongest area in which predictive evidence exists is for differences in moral reasoning between delinquents and non-delinquents. On the other hand there is little support for the idea that postconventional reasoners resist social pressures to commit immoral actions more than those who are predominantly conventional reasoners.

A further source of construct validity for standard issue scoring is that Kohlberg's stages of moral justification can be used as the basis for finding out what sources of information subjects prefer to use in resolving moral issues (Rest et al., 1974; Rest et al., 1978; Rest, 1979, 1983, 1986; Lawrence, 1987). Studies of this kind have all used Rests's Defining Issues Test and have found that there is a tendency for older as well as more intelligent subjects to use sources of information rated in the higher stages. As the DIT shows substantial correlations with the results of moral judgement interviews scored by Kohlberg's methods this provides a source of construct validity if we assume that Kohlberg's stages are associated with types of information sources used to resolve dilemmas. It is not clear that Colby and Kohlberg (1987a)

wish to take this step and thus we must think of this kind of validity as only applying for those who hold Rests's theory that this is so.

Alternative methods of classifying moral justifications have been developed by Williams (1969), Trainer (1982) and Davidson et al. (1983). Two difficulties are shared by all these methods. One is that they do not include level of abstraction in justifications as a dimension, although this is of obvious developmental interest. The second is that, within those types of justification that go beyond offering a rule, none of the above methods copes with the problem that many replies contain more than one kind of justification. Thus the reply that we should do something because 'I think it would be best for everyone concerned' contains both belief and welfare elements and we should either code both or have a rule to decide which predominates. The usefulness of these methods is further limited as Williams (1969) and Davidson et al. (1983) omit substantial categories of justification, while Trainer's (1982) method was exploratory rather than designed for replication.

The most widely used of these methods is that of Davidson et al. (1983) and this has provided some useful evidence about developmental patterns in justifications. These authors found that appeals to authority and punishment avoidance appear to diminish with age for both moral and conventional judgements, while trends for fairness and others' welfare were more various as they differed for the four conditions produced by combinations of the moral-conventional and familiar-unfamiliar variables. This was probably, however, partly due to floor effects, as the only conditions that produced enough responses for reliable trend estimation were familiar moral for fairness and both the moral conditions for others' welfare. Within these conditions, such

responses increased with age, except for others' welfare in the familiar moral condition, which decreased slightly. Although this last effect was not tested statistically, Tisak and Turiel (1984) found that use of others' welfare did decline significantly in the period 6-11 years for familiar moral judgements. However, appeals to fairness, punishment avoidance and authority showed no significant age effect or trend for such judgements in their study. Tisak and Ford (1986) found that there were no significant age trends in the 6-10 year period for any of these kinds of justification used in relation to judgements of inhibitory or active morality, though punishment avoidance did show a tendency to decline with age. It appears from these findings that for familiar moral judgements, use of others' welfare declines in the 6-11 year period, while punishment avoidance and authority justifications decline in some studies, while remaining about the same over the period in others. It may be, as suggested by Trainer (1982), Rest (1983) and Arsenio and Ford (1985) that such disparities are produced by within-domain variation in the situations presented for judgement.

It has also been shown that the distribution of justification types differs between moral, conventional and prudential rule domains (Davidson et al., 1983; Turiel, 1983; Tisak and Turiel, 1984; Arsenio and Ford, 1985), that it differs between familiar and unfamiliar situations (Davidson et al. 1983) and that within the moral domain it differs between instances of inhibitory and active morality (Tisak and Ford, 1986). We can suspect, from the low inter-dilemma reliability of Kohlberg's stages, the work of Trainer (1982), as well as the suggestions of Rest (1983) and Arsenio and Ford (1985), that there are also differences in the balance of justification types elicited by

specific situations and dilemmas within the broad categories of moral problems established by the dimensions of familiarity and inhibitory-active morality.

This brief review indicates that Kohlberg's method of categorising justifications for moral judgements has poor face validity as responses assigned to stages lack unifying conceptual characteristics. This makes interpretation of the construct and predictive validity of the measure problematic. Existing alternative methods of categorising justifications are deficient in a number of respects. Development of an alternative method for categorising justifications that overcomes these deficiencies is of interest as it may permit a more accurate assessment of what Kohlberg's stages measure, as well as leading to a method for assessing moral reasoning with greater face validity.

The object of the first study reported here was to investigate the feasibility and inter-rater reliability of a new method of categorising justifications from a Kohlberg-style moral judgement interview on familiar topics in the range 7 to 15 years, to study developmental trends in justifications when scored in this way, to explore alternative explanations for such observed trends and to assess the influence of the topics presented for judgement on the distribution of justifications .

STUDY 1

Method

Subjects were forty 7 year olds, forty 11 year olds and forty 15 year olds chosen by their teachers as being of about average ability from state primary and secondary schools in Melbourne, Australia. Half in each age group were male and half female.

Two interviewers were used, each interviewing half the children at each age. Each child received the same interview. The style of questioning was Kohlbergian, with subjects being presented with situations requiring moral judgements and asked first what they thought should be done in each situation and then repeatedly pressed as to why that would be right until they were unable to produce further reasons.

The interview schedule contained three parts dealing with: roles; rules and punishment; and personal development and career choice. All these parts dealt with moral judgements as there was always an intrinsic connection between acts suggested by interviewees and situation produced by the acts' effect on human rights or wellbeing (Turiel, 1983; Arsenio

and Ford, 1985). This was even true for those 15 year olds who in the roles section of the interview thought the division of labour within the family should be established by agreement, as here the instrinsic connection was established by the preferences and wellbeing of the father and mother concerned.

Each situation presented was accompanied by a picture depicting the situation in order to make the interview more accessible to the younger children. Pilot studies had shown that the level of abstraction contained in the question influences the level of abstraction contained in justifications. To make the questions suitable for young children, these all asked about the actions of one or two individuals rather than about the actions of groups. The first situation in the roles interview was:'In this picture you can see a mother standing in the window of her house waving father off to work. Do you think it is right for the father to go out to work? Why? Do you think it is right for the mother to stay home and look after the children? Why? Suppose the father didn't want to go out to work. Would it be right for him not to go to work? Why? Suppose the mother didn't want to stay at home to look after the children, but she wanted to go out to work. Would that be right? Why?' This was followed by questions asking if it would be right to make a person go out to work or stay at home if they wanted to do the opposite and questions asking about a family that decided that the mother should always fix the car and about one that decided that the father should always fix the car.

The first situation in the rules and punishment section was based on one used by Piaget (1932), and was:'Here is a little boy/girl who has broken a cup by accident. Do you think that was right? Why? Should they

be punished? Why? What would be the right punishment? Why? Suppose the child broke the cup on purpose. Would that be right? Why? Should they be punished? Why? Suppose they broke eight or ten cups on purpose. Would that be right? Why? What would be the right punishment? Why? ' This was followed by questions asking about a little boy/girl who had a disagreement with his/her mother over tidying his/her room and then by questions about a ten year old who ran out into the road in front of a car while chasing a ball.

The first situation in the personal development and career choice section was: 'A boy/girl wanted to be a farmer/secretary when they grew up. Supposing someone said they couldn't do that. Would it be right for that person to stop them? Why? Is it right to try and stop someone becoming what they want? Why?' This was followed by questions about a boy/girl who spent all their time swimming, neglected their school work and couldn't get a job when they grew up and then by questions about curing a child's shyness.

The following categories were used to code justifications for moral judgements. Categories are arranged along the two dimensions of abstraction and reflection. Each level of reflection contains a number of broad categories each of which comprises two or three levels of abstraction. The level of abstraction is determined by the reference group used in justifications, which may refer to an individual, to local groups such as family or friends or to wider groups, such as Australians or people in general. The three levels of reflection are descriptive (rules and rights), naturalistic (based on what people tend to do or think), theoretical (providing a theoretical reason for rules).

Descriptive categories comprise only one broad category, which includes all justifications providing us with a general rule that can be used to deduce the judgement or rule being justified. This contains three levels of abstraction: rules applying to local situations; rules applying to large groups of people or to people in general; rules of the latter type that are explicitly said to be 'rights'.

Naturalistic categories comprise three broad categories: justifications based on the idea that certain people think something is right (Belief); those based on the idea that people will inevitably behave in a certain way and this makes it right (Behavior); those based on the idea that by obeying rules we can avoid punishment (Punishment). Each of these categories can refer either to an individual, local groups or wider groups.

Theoretical categories comprise three broad categories, two focussing on the benefits of actions for others, one on religious justifications. The first benefit category (Interest) justifies judgements or rules by the argument that they promote the interests or abilities of an individual, a local group or a wider group (including people in general). The second benefit category (Social Harmony) justifies a judgement or rule by reference to the need to maintain social order or harmony. The reference may be either to the harmony of a local group or to that of a wider group (e.g. 'society'). The religious category (Religious) justifies a judgement or rule by reference to the idea it is divinely sanctioned or inspired. This may be used to justify the belief of an individual, of a local group or of a wider group such as a church or nation.

Because justifications often contain elements from more than one of the broad categories the method of categorisation described above has either to be applied so that each justification can score in more than one category or using precedence rules that establish which element determines the categorisation of the reply. The latter is used here, as the former method complicates both the presentation of results as well as the scorer's task.

Any item containing elements from more than one level of reflection is scored according to the element at the highest level. For example, if the interviewee says 'I think people might trip over' as a justification for the judgement that a child should tidy their room, then the interest element in 'people might trip over' overrides the belief element in 'I think'. Cases where there is more than one element at the highest level of reflection found in the item are rare and have to be decided by the relative emphasis accorded to each.

Results and Discussion

Each interview was tape recorded and the transcript of the recording marked up by the first investigator to show replies justifying a moral judgement or rule. This meant, in particular, that justifications for the magnitudes of punishments were not scored. Each transcript was scored independently by two raters who were trained in the scoring method using sample transcripts and a scoring manual that contained both the definitions of categories given above and examples of replies falling into the categories taken from a pilot study using the same interview.

Findings about justifications for judgements and rules will be reported separately from those on the content of judgements and rules, following the suggestions of Davidson et al. (1983) and Kaplan (in press) that these aspects should be clearly separated. The percentages of replies falling within broad categories of justification are relatively numerous and stable and are shown in Table 1. The percentages of replies falling into some of the subcategories formed by the combination of level of abstraction and broad category is quite small and unreliable. Table 2 only breaks down the three categories that attracted a large enough percentage of replies to make such a breakdown meaningful in a study of this size. Percentages were reasonably constant across interviewers and raters for the categories and subcategories reported in both tables and so both show average percentages for the four raters.

- Tables 1 and 2 -

Inter-rater agreement was 84 and 86 percent for the two interviewers.

Univariate two-way analysis of variance was carried out on the percentage of a subject's scores that fell within each of the categories listed in Table 1, except Harmony and Relgious. Age and interview topic were the independent variables, with repeated measures on the interview topic factor. This analysis showed that the percentage of an interviewee's justifications falling into both Descriptive and Interest categories was influenced by the interview topic ($F(2,57) = 3.38$ and $F(2,57) = 3.26$, $ps < .05$). The roles section of the interview attracted the greatest percentage of Interest justifications, while the rules and

punishment section attracted the greatest percentage of Descriptive justifications. Interest justifications did not decline with age for the roles section. As Davidson et al. (1983) found Interest type justifications markedly reduced for conventional judgements, this supports the view that roles questions were seen as relating to moral issues even by the older interviewees. These findings support the suggestions of Trainer (1982), Rest (1983) and Arsenio and Ford (1985) that there are appreciable within-domain variations in the balance of types of justification.

The analysis of variance also showed that age had a significant influence on the percentage of responses falling into the Descriptive and Interest categories ($F(2,114) = 3.42$ and $F(2,114) = 3.15$, $ps < .05$). It can be seen from Table 1 that these effects were most marked in the 7 to 11 interval with Descriptive increasing and Interest declining in this period, while percentages within these categories remained roughly constant in the 11 to 15 interval. The differences between the 7 and 11 year groups in the percentage of responses falling into Descriptve and Interest categories were tested using two-tailed t-tests and found to be significant ($t(78) = 1.89$, $t(78) = 1.95$, $ps < .05$). The finding that Interest justifications for familiar moral judgements decline in the 7-11 year period is consistent with the findings of Davidson et al. (1983), Tisak and Turiel (1984) and Tisak and Ford (1986) as their others' welfare category is a substantial component of this type of justification. There were no significant interaction effects.

Pilot studies taken with the findings of Langford and D'Cruz (in press) had suggested three developmental trends involving shifts in the level of abstraction of justifications: 1. Within the Descriptive

category there would be a shift from concrete to abstract responses with increasing age; 2. Within the Belief category the ratio of local to individual responses would peak at around 15/16 (and decline thereafter, though this would not be observed in the present study); 3. Within the Interest category there would be a shift from concrete to abstract responses with increasing age. Trends 1, 2 and 3 imply that the average level of abstraction will follow the order 1,2,3 (least to most) with increasing age. This suggestion can be tested against the the null hypothesis that ordering is random by computing the probability that both interviewers would exhibit the predicted ranking under the null hypothesis. As both interviewers showed all three trends, after averaging over raters, the null hypothesis may be rejected in every case, ps .05.

Because of the particular interest that justifications based on the beliefs or wishes of authority have for understanding the thinking of younger children, the categories of Individual and Local belief were divided into the following subcategories. Individual Belief was divided into Personal Belief (belief of the interviewee); Belief of Authority (usually parents or teachers); and Justification for the Belief or Wishes of Authority (usually the claims that they know more, are older or they look after you), while Local Belief was divided into Belief of Equals (e.g. friends, brothers and sisters); Belief of Authorities (usually parents, police or teachers); or Justifications for the Beliefs or Wishes of Authorities. It was expected that the ratio of Personal Belief replies to Authority Belief replies and the ratio of Belief of Equals to Belief of Authorities would increase with age. These trends were both present in the average ratings for both interviewers and thus

the null hypotheses could be rejected in regard to these trends, ps .05. In the first case the ratio increased from about 3 to 2 to about 4 to 1, in the second it increased from about 1 to 3 to about 2 to 1.

The data on percentages of justifications falling in the various categories at different ages, taken with the developmental trends, suggest that such percentages and trends are better explained in terms of changes in the focus of interviewees' social interests and shifts in their use of language rather than in terms of their cognitive limitations.

Trends in descriptive justifications seem to stem chiefly from a shift in subjects' use of language. Everyday experience as well as the data in Table 2 show that most 7 year olds can quite readily conceive of social rules that have a broad range of application (Wider Descriptive). The balance between local rules and rules with wider application does not change much with age, when we take into account that Rights Descriptive replies are at the same level of generality as Wider Descriptive. By age 15 there is an increased tendency to call wider rules 'rights'. This seems to be largely a socially acquired method of underlining the importance of certain rules by describing them as rights, rather than a fundamental change in level of thinking. The question that produced most examples of those universal human rights that Kohlberg thought of as rights in the stage 5 sense was one on personal development and career choice asking if it is right for another person to prevent a child from taking up the profession of their choice. However, this right to choose was asserted by 90 per cent of the 7 year olds, as opposed to 100 per cent of 15 year olds. At the same time, the second group sometimes used the term 'right' to describe this belief,

while the first never did. This contrasts with the issue of tidying one's room, where 93 percent of 7 year olds said that a child should tidy their room when asked by their mother, while only 38 per cent of 11 year olds and 35 per cent of fifteen year olds did so. Here older children quite often discussed the issue of the child's 'rights' versus the mother's 'rights', while younger children never did. In general the fact that a rule is described as a 'right' does not make it a different kind of rule.

The tendency for Belief justifications to shift towards validation by local and wider groups rather than the individual in the 7 to 10 year period is again unlikely to result from increased ability to manage conceptual difficulty as most 7 year olds are aware that families and groups of friends as well as wider groups share beliefs. Their tendency to cite the beliefs of individuals is more likely to stem from social egocentrism and an interest in authority figures. This is later counterbalanced by a focus on the peer group and the family conceived as an association of equals, as well as on the wider society.

The tendency for interest justifications to shift focus from the individual to the local group is also unlikely to result from purely cognitive changes as most 7 year olds can readily conceive of the interests of local groups; this shift in abstraction level is more plausibly explained by greater social interest in the peer group and other local groups.

There were three notable developmental trends in the content of judgements and rules. Descriptive and inferential statistics on these will not be given as they are familiar tendencies from other studies, but a brief summary of findings here gives a rounder picture of overall

findings and is relevant to later discussion. The first was that in the roles section of the interview younger children thought of the duties attached to roles as quite stereotyped and unalterable, while older children became less stereotyped and began to think of work roles within the family as something arranged by agreement. The second trend was that older children showed a greater ability to discount an unintentional transgression, although most 7 year olds did make some distinction between unintentional and intentional transgressions. This is consistent with studies showing that older children give greater weight to intentionality as opposed to magnitude of consequences than younger children (for reviews see Lickona, 1976; Karniol, 1978; Keasey, 1978; Tomlinson, 1980; Ferguson and Rule, 1982; Shultz et al. , 1986). The third trend was that replies to two questions showed a developmental shift from paternalism and the rights of authority to autonomism and the right of the child to choose. Although Kohlberg has often pointed to the growth of autonomy as an aspect of moral development, his method of indexing this through moral stages tends to confound this trend with others. This point will be developed more fully in the discussion of Study 2.

STUDY 2

The second study involved the administration of Kohlberg's Form A interview schedule to subjects in the age range 7 to 21 years, with a view to comparing the Colby and Kohlberg (1987b) scoring method with the method described in the first study. An attempt was made to score justifications from the first study using Colby and Kohlberg's (1987a) stages, but this proved unsatisfactory due to the difficulty of

proceeding from stage definitions to scoring individual justifications that were not in the (1987b) manual.

Inspection of the way in which the new method of categorising justifications scores the criterion judgements listed in Colby and Kohlberg's (1987b) manual for the Form A interview showed a consistent pattern of relations between the two methods. This was that categories and subcategories from the method of Study 1 spanned criterion judgements from a band of Kohlberg stages. More specifically, criterion judgements classed as Wider Descriptive, Individual Belief or Religious belonged to stages 1 to 4/5; those as Individual Interest, Local Interest, Behavior or Punishment to stages 1 to 3/4; those as Local Descriptive to stages 2 to 3/4; those as Local Harmony to stages 3-4; those as Wider Interest to stages 2-5; those as Wider Harmony to stages 3-5; those as Rights Descriptive to stage 5. About one percent of criterion judgements were unclassified. Of the remainder only five percent did not conform to the above generalisations. It is also the case that within the Behavior, Punishment and Religious categories, criterion judgements scored as more abstract by the alternative coding method also tended to belong to higher stages. As these categories are relatively infrequent this will be ignored in the empirical analysis. (The comments in this paragraph are based on figures collected omitting punishment issue criterion judgements as many of the items the scorer is directed to score are not justifications for moral judgements but for the pragmatics of punishment.)

Inspection of the descriptions of criterion judgements in the manual revealed that within the Local Descriptive, Wider Descriptive, Individual Belief, Wider Belief and Individual Interest categories, but

not within the other categories and subcategories listed in Tables 1 and 2, there is a tendency for justifications scored at higher stages to specify a content for the judgement or rule described that is more likely to be found among older respondents than among younger. This is chiefly due to the assignment of replies asserting the right to make autonomous behavioral choices and the right to equality with parents to higher stages. It was therefore predicted that within these categories and subcategories there would be a tendency for replies to shift towards a higher average stage rating among older respondents, but that this would not be observed for the other categories and subcategories listed in Tables 1 and 2.

Method

Subjects were thirty two 7/8 year olds, thirty two 11/12 year olds, thirty two 15/16 year olds and thirty two 19-21 year olds chosen by their teachers as being of about average ability from state primary and secondary schools and a university in Melbourne, Australia. Half in each age group were male and half female. The university students were average for an above average group and this may have exaggerated some of the developmental trends in the 16 to 21 year age range when compared with the general population.

Two interviewers were used, each interviewing half the subjects at each age. Each subject received Kohlberg's Form A interview in the form in which it appears in Colby and Kohlberg (1987b), omitting the optional starred questions.

Results and Discussion

Each interviewer's transcripts were scored by one expert rater using both the method of Colby and Kohlberg (1987b) and that used in the first study. The assignment of justifications by both methods is shown in Tables 3-6. These show that the variation in stage scores is greater at each age than found by Colby and Kohlberg (1987a) for their American sample. This is probably largely caused by their use of the rule that if less than ten percent of a subject's replies fall within a given stage then these are treated as error and not scored (p. 90).

- Tables 3 to 6 -

Although five percent of criterion judgements failed to fit the suggested generalisations about stage bands for the categories and subcategories of the alternative system, less than one percent of replies at each age failed to conform to these generalisations.

Predictions regarding the tendency for replies within some of the alternative categories to show a trend towards a higher average stage score with increasing age were tested by computing the average stage score within the category or subcategory at each age. Under the null hypothesis that average stage score is independent of age, the observation that there is a perfect correlation between average stage score and age has a probablity of 1 in 24. On this basis, the trends for the Wider Descriptive and Individual Belief categories are significant, $ps \leq .05$. There are signs of these trends for the Local Descriptive, Wider Belief and Individual Interest categories, but they are not significant due to slight reversals.

In interpreting these findings it should be mentioned that the assignment of the replies of older subjects to higher Kohlberg stages within the categories and subcategories mentioned may not be solely due to differences in the content of rules and judgements but to aspects of the form of justifications not captured by the alternative coding system. While this could be so, there is no doubt that a large number of replies are assigned to higher stages within the alternative categories because of the content of the rules and judgements contained in them. This is damaging evidence against Colby and Kohlberg's (1987b) method of categorisation as this claims to concentrate on forms of justification. It is difficult to know what further aspects of form of justification might contribute to the observed effects, due to the lack of clear conceptual definitions for the stages. If such further aspects were shown to be of significance, these should be incorporated within a categorisation that concentrates solely on forms of justification, rather than considered as support for Colby and Kohlberg's method, which involves serious confounding of the content of judgements and rules and forms of justification.

Tables 3-6 show the classic tendencies for modal Kohlberg stage score to increase with age and for stages closer to the modal score to attract more scores than those further removed. These findings can be explained as a joint product of developmental changes in subjects' preferred forms of justification (as indicated by the alternative coding method) and changes in the content of rules and judgements that result in older subjects' replies being assigned to higher stages. The forms of justification described here as Local Interest and Punishment are more frequent among younger children and result in responses being assigned

to lower Kohlberg stages. The form described as Local Harmony declines in frequency somewhat later than the forms just mentioned and results in assignment to intermediate Kohlberg stages. The forms described as Wider Interest, Wider Harmony and Rights Descriptive are more common among older children and result in assignment to higher stages. At the same time there is a shift within the forms discussed in the previous two paragraphs towards higher stages with increasing age. This appears to be due to age related changes in the content of rules and judgements within these categories.

It is even harder to explain the data in Tables 3-6 by means of cognitive limitation arguments than that in Table 2 because average level of abstraction in the Belief category declines in the period 16-21 years, thus replicating the findings of Langford and D'Cruz (in press). This can be explained by the social focus approach as resulting from a shift away from interest in the beliefs of others towards self reliance.

According to the approach suggested here, the growth of autonomous moral reasoning is indexed by a change in the content of rules towards the assertion of autonomous behavioral choice. It is also indexed by a shift away from the justification of rules by authority towards their justification by personal belief. This second trend is however a complex one, as while the ratio of personal to authority belief justifications in Study 1 increased in the period from 7 to 15 years, in this same period the ratio of Personal Belief justifications to all justifications actually declines somewhat from 6 to 4 percent because of the tendency to give more Local and Wider Belief justifications. In the second study the percentage of Personal Belief to all justifications also declined up to 15 years, but then increased slightly in the period from 16 to 21. As

this second trend was also observed for moral judgement interviews not contained in Kohlberg's Moral Judgement Interview by Langford and D'Cruz (in press) it probably has reasonable generality. This way of describing the shift towards autonomous reasoning is more explicit than that of Colby and Kohlberg (1987b), as it avoids both confounding the content of judgements and rules with forms of justification and confounding the trend towards autonomy within each of these areas with other trends.

One outcome of this alternative way of indexing the growth towards autonomy is to highlight that such growth involves a relative increase in autonomous thinking, rather than its unreserved adoption either in the content of rules or in justification. On some issues even 7 year olds use the rule that everyone should decide for themselves (e.g. in career choice), on some issues there is a shift towards autonomy (e.g. tidying one's room), while on others (e.g. children endangering themselves) it is rarely asserted by adults. At the same time, justifications based on personal belief always remain a small minority of all justifications.

These findings may also throw new light on the issue of the construct and predictive validity of Kohlberg's stage sequence. They suggest a possible reinterpretation of findings from the Defining Issues Test that show that information sources corresponding to justification types belonging to Kohlberg's more advanced stages are preferred by older subjects. This may be mainly because of a shift in the social focus of the subject's interests and social relationships rather than primarily because of improvements in cognitive maturity. The finding that Colby and Kohlberg's (1987b) method of assigning replies to stages confounds the content of rules with form of justification suggests it

may be profitable in future studies of the DIT and of behavioral prediction to clearly separate these aspects.

CONCLUSIONS

The two studies reported show that a new method of categorising justifications for moral judgements has acceptable levels of inter-rater reliability when applied to data from moral judgement interviews. This method permits assignment of justifications to categories on the basis of category definition rather than by the use of *ad hoc* assignment of criterion judgements to stages as occurs in Colby and Kohlberg (1987b). It also permits a clearer distinction between the form of justifications and the content of rules and judgements than Kohlberg's methods. This makes possible a clearer view of the growth towards autonomous moral reasoning than Kohlberg's stage progression. The method also overcomes some of the limitations of existing non-Kohlbergian methods of categorising justifications. Use of the method reveals a pattern of developmental trends that is better accounted for by a social focus rather than a cognitive limitation view of the development of moral reasoning. It is suggested that future studies of the DIT, as well as of behavioral prediction from indices of moral reasoning, could be assisted by application of the scoring method described.

REFERENCES

ARSENIO, W.F., & FORD, M.E (1985). The role of affective information in social-cognitive development; children's differentiation of moral and conventional events. Merrill-Palmer Quarterly, 31, 1-17.

BLASI, A. (1980). Bridging moral cognition and moral action: a critical review of the literature. Psychological Bulletin, 88, 1-45.

BROUGHTON, J.M. (1975). The cognitive-developmental approach to morality: a reply to Kurtines and Greif. Journal of Moral Education, 8, 81-96.

COLBY, A., & KOHLBERG, L. (1987a,b). The measurement of moral judgement, Volumes 1 and 2. Cambridge: Cambridge University Press.

CORTESE, A.J. (1984). Standard issue scoring of moral reasoning: a critique. Merrill-Palmer Quarterly, 30, 227-246.

CRITTENDEN, B.S. (1978). Bearings in moral education. Melbourne: ACER.

DAVIDSON, P., TURIEL, E., & BLACK, A. (1983) The effect of stimulus familiarity on the use of criteria and justifications in children's social reasoning. British Journal of Developmental Psychology, 1, 49-65.

DAVISON, M.L. (1979). The internal structure and the psychometric properties of the Defining Issues Test. In J. Rest (ed.) Development

in judging moral issues (pp.223-245). Minneapolis: University of Minnesota Press.

FERGUSON, T.J.,& RULE, B.G. (1982). Influence of inferential set, outcome intent, and outcome severity on children's moral judgements. Developmental Psychology, 18, 843-851.

GIBBS, J.C., WIDAMAN, K.F., & COLBY, A. (1982). Construction and validation of a simplified, group administerable equivalent of the moral judgement interview. Child Development, 53, 895-910.

GILLIGAN, C. (1982). In a different voice: Essays on psychological theory and women's development. Cambridge: Harvard University Press.

KAPLAN, M.F. (in press) Information integration in moral reasoning: conceptual and methodological implications. In N.H. Eisenberg (Ed.), Social and moral values: individual amnd societal perspectives. Hillsdale: Lawrence Erlbaum.

KARNIOL, R. (1978). Children's use of intention cues in evaluating behavior. Psychological Bulletin, 85, 76-85.

KEASEY, C.B. (1978). Children's developing awareness and usage of intentionality and motives. In C.B. Keasey (Ed.), Nebraska Symposium on Motivation (pp. 219-260). Lincoln: University of Nebraska Press.

KOHLBERG, L., & KRAMER, R. (1969). Continuities and discontinuities in childhood and adult moral development. Human Development, 12, 93-120.

KUHN, D. (1976). Short-term longitudinal evidence for the sequentiality of Kohlberg's early stages of moral development. Developmental Psychology, 12, 162-166.

LAWRENCE, J.A. (1987). Verbal processing of the defining issues test by principled and non-principled moral reasoners. Journal of Moral Education, 16, 117-130.

LANGFORD, P.E., & D'CRUZ, J.V. (in press). A new way of scoring moral judgement interviews. Journal of Moral Education, 19.

LICKONA, T. (Ed.) (1976). Moral development and moral behaviour: Theory, research and social issues. New York: Holt, Rinehart & Winston.

PIAGET, J. (1932). The moral judgement of the child. London: Routledge and Kegan Paul.

REST, J.R. (1979). Development in judging moral issues. Minneapolis: University of Minnesota Press.

REST, J.R. (1983). Morality. In J.H. Flavell & E. Markman (Eds.), Handbook of child psychology: Vol. 3. Cognitive development (pp. 556-629). New York: Wiley.

REST, J.R. (1986). Moral development: advances in research and theory. New York: Praeger.

REST, J.R., COOPER, D., CODER, R., MASANZ, J., & ANDERSON, D. (1974). Judging the important issues in moral dilemmas - an objective test of development. Developmental Psychology, 10, 491-501.

REST, J.R., DAVISON, M., & ROBBINS, S. (1978). Age trends in judging moral issues: a review of cross-sectional, longitudinal and sequential studies of the Defining Issues Test. Child Development, 49, 263-279.

SHULTZ, T.R., WRIGHT, K., & SCHEIFER, M. (1986). Assignment of moral responsibility and punishment. Child Development, 57, 177-184.

SIEGAL, M. (1980). Kohlberg versus Piaget: to what extent has one theory eclipsed the other? Merrill-Palmer Quarterly, 26, 285-297.

SIMPSON, E.L. (1974). Moral development research: a case study of scientific cultural bias. Human Development, 17, 81-106.

TISAK, M.S., & FORD, M.E. (1986) Children's conceptions of interpersonal events. Merrill-Palmer Quarterly, 32, 291-306.

TISAK, M.S., & TURIEL, E. (1984). Children's conceptions of moral and prudential rules. Child Development, 55, 1030-1039.

TOMLINSON, P. (1980). Moral judgement and moral psychology: Piaget, Kohlberg and beyond. In S. & C. Modgil (Eds), Toward a theory of psychological development (pp. 303-366). Slough: NFER.

TRAINER, F.E. (1977). A critical analysis of Kohlberg's contribution to the study of moral thought. Journal for the Theory of Social Behaviour, 1, 41-63.

TRAINER, F.E. (1982) Dimensions of moral thought. Sydney: University of New South Wales Press.

TURIEL, E. (1983) The development of social knowledge. Cambridge: Cambridge University Press.

WILLIAMS, N. (1969). Children's moral thought. Moral Education, 1, 3-12.

YUSSEN, S.R. (1977). Characteristics of moral dilemmas written by adolescents. Developmental Psychology, 13, 162-163.